Convergence Rate for Quasi-Newton Methods

Consider the quasi-Newton iteration

$$x_{k+1} = x_k - \frac{F(x_k)}{A_k}, \quad k = 0, 1, \ldots. \tag{1}$$

Assume (i) $F : R \to R$ is twice continuously differentiable; (ii) $F(x^*) = 0$; (iii) $F'(x^*) \neq 0$; (iv) $x_k - x^* \to 0$ as $k \to \infty$; and (v)

$$F'(x_k) - A_k \to 0 \quad \text{as} \quad k \to \infty. \tag{2}$$

Note that we have assumed the iteration convergences. With some additional assumptions (e.g., given that x_0 is sufficiently close to x^* and given specific bounds in eqn (2)) one can actually prove convergence using the contraction mapping theorem.

Main Result. Under the above assumptions, the iteration (1) converges to x^* at a *super-linear rate.* This means

$$|x_{k+1} - x^*| \leq c_k |x_k - x^*| \tag{3}$$

for some sequence of positive numbers c_k for which $\lim_{k\to\infty} c_k = 0$.

Proof: Define the error $e_k = x_k - x^*$. Because of assumption (ii),

$$x^* = x^* - \frac{F(x^*)}{A_k}. \tag{4}$$

Subtracting eqn (4) from eqn (1) gives

$$e_{k+1} = e_k + \frac{1}{A_k}(F(x^*) - F(x_k)). \tag{5}$$

Applying a Taylor expansion to $F(x)$ about $x = x_k$ and letting $R(x_k, x^*)$ denote the Taylor remainder term, we get

$$\begin{aligned} F(x^*) &= F(x_k) + F'(x_k)(x^* - x_k) + R(x_k, x^*) \\ &= F(x_k) + (A_k + F'(x_k) - A_k)(x^* - x_k) + R(x_k, x^*) \\ &= F(x_k) - A_k e_k + (A_k - F'(x_k))e_k + R(x_k, x^*) \end{aligned} \tag{6}$$

Substituting (6) into (5) and doing some algebra gives

$$e_{k+1} = \frac{A_k - F'(x_k)}{A_k} e_k + \frac{R(x_k, x^*)}{A_k} \tag{7}$$

We next apply the triangle inequality, $|a + b| \leq |a| + |b|$ to to get

$$\begin{aligned} |e_{k+1}| &\leq \frac{|A_k - F'(x_k)|}{|A_k|}|e_k| + \frac{|R(x_k, x^*)|}{|A_k|} \\ &\leq \frac{|A_k - F'(x_k)|}{A_{\min}}|e_k| + \frac{|R(x_k, x^*)|}{A_{\min}}, \end{aligned} \tag{8}$$

where $|A_k| \geq A_{\min} > 0$ for k sufficiently large. That such an $A_{\min}$ exists follows from assumptions (iii) and (v).

By assumption (i), the Taylor remainder term $R(x_k, x^*) = \mathcal{O}(e_k^2)$. This means

$$|R(x_k, x^*)| \leq C|e_k|^2 \tag{9}$$

for some constant $C > 0$. By substituting (9) into (8), we get for k sufficiently large that

$$\begin{aligned} |e_{k+1}| &\leq \frac{|A_k - F'(x_k)|}{A_{\min}}|e_k| + \frac{C|e_k|^2}{A_{\min}} \\ &= \left(\frac{|A_k - F'(x_k)|}{A_{\min}} + \frac{C|e_k|}{A_{\min}}\right)|e_k|. \end{aligned} \tag{10}$$

The quantity inside the parentheses becomes the c_k in the superlinear convergence definition (3). That $c_k \longrightarrow 0$ as $k \longrightarrow \infty$ is a consequence of assumptions (iv) and (v).

Note that for Newton's method, $A_k = F'(x_k)$. Then from (10) we get a quadratic convergence rate,

$$|e_{k+1}| \leq \frac{C}{\mathrm{Amin}}|e_k|^2.$$